(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 315 356**

(51) Int. Cl.:
G06F 19/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **02718498 .5**
(96) Fecha de presentación : **26.03.2002**
(97) Número de publicación de la solicitud: **1488364**
(97) Fecha de publicación de la solicitud: **22.12.2004**

(54) Título: **Procedimiento y sistema para construir modelos tridimensionales de estructuras moleculares.**

(45) Fecha de publicación de la mención BOPI:
01.04.2009

(45) Fecha de la publicación del folleto de la patente:
01.04.2009

(73) Titular/es:
Council of Scientific and Industrial Research
Rafi Marg
New Delhi 110 001, IN

(72) Inventor/es: **Gautham, N. y**
Vengadesan, K.

(74) Agente: **Carpintero López, Mario**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Procedimiento y sistema para construir modelos tridimensionales de estructuras moleculares.

Campo de la invención

La presente invención se refiere a un procedimiento para construir modelos tridimensionales de estructuras moleculares correspondientes al menor valor de la función de energía potencial a partir del conocimiento de sus estructuras químicas. La presente invención especifica un procedimiento que se puede usar para construir un modelo de la estructura tridimensional de cualquier molécula que incluye, pero sin limitaciones, péptidos, proteínas, nucleótidos, ácidos nucleicos, antibióticos, sacáridos, etc. El sistema usa Cuadrados Latinos Mutuamente Ortogonales para buscar a través del espacio de conformación para obtener la mejor conformación. El sistema también se puede usar para obtener el óptimo de cualquier función, y no sólo las relacionadas con la estructura biomolecular. El procedimiento se incorpora en una computadora y este sistema forma la presente invención.

Antecedentes de la invención

La patente de los Estados Unidos N.º 5.884.230 de Srinivasan *et al*., intitulada "Procedimiento y sistema para modelado de proteínas" describe un procedimiento en un sistema de computación para modelar una estructura tridimensional de una proteína modelo en la cual el modelado se basa en una estructura tridimensional de una plantilla proteica plantilla proteica y un alineamiento de la secuencia de aminoácidos de la proteína modelo y de la plantilla proteica. Para cada aminoácido en la proteína modelo, cuando la plantilla proteica tiene un aminoácido alineado el aminoácido de la proteína modelo, la posición del átomo del esqueleto del aminoácido alineado con el aminoácido de la proteína modelo se establece sobre la base de la posición de un átomo del esqueleto topológicamente equivalente en el aminoácido alineado de la plantilla proteica. Sin embargo, este procedimiento sólo es operativo cuando ya se conoce la estructura de una proteína modelo. En consecuencia, no es aplicable a moléculas más pequeñas tales como oligopéptidos, o a moléculas que son químicamente diferentes, tales como ácidos nucleicos, sacáridos, antibióticos, y otras moléculas orgánicas. Incluso en el caso de una estructura proteica, el procedimiento no es aplicable a la determinación de una estructura proteica *ab initio*, en otras palabras, el procedimiento no se puede aplicar si no hay una estructura modelo de la cual comenzar.

La patente de los Estados Unidos N.º 6.188.956 de Mayo, *et al*., intitulada "Aparato y procedimiento para el diseño automático de proteínas" se refiere a aparatos y procedimientos para el diseño y la optimización cuantitativos de proteínas. El procedimiento recibe una estructura de esqueleto proteico conocido y construye las conformaciones de las cadenas laterales usando el procedimiento de Eliminación de Extremos sin Salida. Igual que antes, este procedimiento también es operativo solamente cuando ya se conoce la estructura de una proteína modelo. En consecuencia, no es aplicable a moléculas más pequeñas tales como oligopéptidos, o a moléculas que son químicamente diferentes, tales como ácidos nucleicos, sacáridos, antibióticos, y otras moléculas orgánicas. Incluso en el caso de estructuras proteicas, el procedimiento no es aplicable para la determinación *ab initio* de estructuras proteicas; en otras palabras, si no hay una estructura modelo a partir de la cual comenzar, el procedimiento no se puede aplicar, dado que requiere una conformación ya conocida del esqueleto proteico.

"*Combinatorial Mathematics*", de H. J. Ryster, un libro de texto de matemáticas combinatoria provee una definición de Cuadrados Latinos Mutuamente Ortogonales, junto con una ecuación para construirlos. Sin embargo, no hay explicaciones sobre cómo se debe incorporar la ecuación en un programa de computación. Tampoco hay mención sobre cualquier aplicación de los Cuadrados Latinos Mutuamente Ortogonales (MOLS) para hallar el mejor valor de cualquier función, incluso la función de energía potencial. Tampoco hay mención sobre cualquier aplicación para la determinación de la estructura tridimensional de una estructura molecular. Además, sólo describe cómo distribuir un conjunto de símbolos con la finalidad de formar un conjunto de MOLS, pero no describe un procedimiento para reemplazar los símbolos con los valores de los parámetros que definen cualquier función. Tampoco describe cómo se deben analizar los cálculos después de la construcción de los MOLS.

"*Introduction to Combinatorial Mathematics*", de C. L. Liu, es otro libro de texto de referencia para la matemática combinatoria que provee una definición de Cuadrados Latinos Mutuamente Ortogonales, junto con una ecuación para construirlos. Sin embargo nuevamente no hay ninguna explicación sobre cómo se debe incorporar la ecuación en un programa de computación. Tampoco hay mención sobre cualquier aplicación de los Cuadrados Latinos Mutuamente Ortogonales (MOLS) para hallar el mejor valor de cualquier función, incluso la función de energía potencial. Tampoco hay mención sobre cualquier aplicación para la determinación de la estructura tridimensional de una estructura molecular. Además, sólo describe cómo distribuir un conjunto de símbolos con la finalidad de formar un conjunto de MOLS, pero no describe un procedimiento para reemplazar los símbolos con los valores de los parámetros que definen cualquier función. Tampoco describe cómo se deben analizar los cálculos después de la construcción de los MOLS.

"*Experimental Designs and Its Statistical Basis*", de D. J. Finney es un libro de texto sobre el diseño experimental que provee un procedimiento para usar Cuadrados Latinos Mutuamente Ortogonales en experimentos agrícolas y farmacéuticos. Sin embargo, allí no hay enseñanzas sobre cómo construir los MOLS, ni se especifica cualquier programa de computación para hacerlo, ni se menciona cualquier aplicación para la determinación de la estructura molecular o para hallar el mejor valor de una función. Además, los análisis de los cálculos realizados en cada subcuadrado, tal como se establecen en este texto son más complicados que los usados en la presente invención.

"*Genetis Algorithms and Protein Folding*" de S. Schulze-Kremer, Procedimientos en Biología Molecular (D Webster, ed.) 2000 describe la aplicación de algoritmos genéticos en la predicción de la estructura de proteínas. Sin embargo, el procedimiento describe no realiza la predicción *ab initio* de las estructuras. Es decir, si no hay información previa sobre la estructura secundaria, el procedimiento no funciona. Además, el procedimiento no usa MOLS en cualquier forma en los cálculos, y en consecuencia es diferente de la presente invención. "*Global Search for Optimal Biomolecular Structures using Mutually Orthogonal Latin Cuadrados*" de N. Gautham y Z. A. Rafi, Current Science, 1992 es un informe muy preliminar sobre el posible uso de MOLS en la predicción de estructuras. El procedimiento, sin embargo, era muy crudo y no formulado adecuadamente. No incorporaba cualquier ecuación para predecir MOLS, y no especificaba cómo se deben relacionar los símbolos de MOLS con los valores de las variables. No especificaba en detalle el procedimiento usado para analizar los valores calculados en los subcuadrados. No se aplicaba a pentapéptidos.

"*Revised algorithms for build-up procedimiento for predicting protein conformations by energt minimization*" de K. D. Gibson y H. A. Scheraga, *Journal of Computational Chemistry*, 1987 describe un "Procedimiento de Construcción" para predecir conformaciones proteicas. Este procedimiento realiza una búsqueda truncada dependiente de la dominancia de las interacciones a corto plazo. En consecuencia, busca el mínimo local de fragmentos cortos mediante un exhaustivo procedimiento de minimización de la energía. La cantidad de conformaciones de los fragmentos cuya energía debe ser minimizada y almacenada en cada etapa aumenta exponencialmente. El procedimiento se torna inmanejable para cadenas de polipéptidos de aproximadamente 20 residuos de aminoácidos. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento. "*On the multiple-minima problem in the conformational analysis of polypeptides I. Backbone degrees of freedom for a perturbed a-helix*" de L. Piela y H. A. Scheraga, *Biopolymers*, 1987 describe el uso de "Campo electrostático autoconsistente", sobre la base del concepto de que los dipolos del grupo peptídico en conformación nativa deben tener orientaciones aproximadamente óptimas en el campo eléctrico generado por la totalidad de la molécula y su solvente circundante. El procedimiento no usa todos los demás términos de energía. En consecuencia sólo es específico para el campo electrostático. No se puede aplicar a moléculas pequeñas tales como péptidos, etc. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma en el procedimiento.

"*Monte Carlo-Minimization approach to the multiple-minima problem in protein holding*" de Z. Li y H. A. Scheraga, Proceedings of the National Academy of Sciences, 1987 describe un abordaje estocástico para la optimización global de polipéptidos y proteínas que combina la fuerza del procedimiento Metropolis Monte Carlo en la optimización combinatoria global con la de la minimización de la energía convencional para hallar los mínimos locales. Este no es un procedimiento de búsqueda determinística y en consecuencia no asegura la mejor estructura cada vez. También es muy computacionalmente intensivo y, en consecuencia, muy onerosa en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma en el procedimiento y en consecuencia este procedimiento es diferente de la presente invención. "*New developments of the electrostatically driven Monte Carlo method - Test on the membrane bound portion of melittin*" de D. R. Ripoll, A. Liwo y H. A. Scheraga, Biopolymers, 1998 describe un procedimiento iterativo para la búsqueda de la superficie conformacional de los polipéptidos. Este no es un procedimiento de búsqueda determinístico y en consecuencia no asegura la mejor estructura cada vez. También es muy computacionalmente intensivo y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*The multiple-minima problem in the conformational analysis of molecule-Deformation of the potencial energy hipersurface by the diffusion equation method*", de L. Piela, J. Kostrowicki y H. A. Scheraga, Journal of Physical Chemistry, 1989 describe el uso del concepto básico del "Procedimiento de ecuación de difusión" para deformar la función multivariable que representa la energía potencial de manera tal que causa la desaparición gradual de las fuentes poco profundas, mientras que otras fuentes potenciales crecen a su expensa. El procedimiento no se aplica fácilmente en la práctica a estructuras complicadas. Este no es un procedimiento de búsqueda determinístico y en consecuencia no asegura la mejor estructura cada vez. También es muy computacionalmente intensivo y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*Molecular dynamics on deformed potencial energy hypersurfaces*" de J. Pillardy y L. Piela, Journal of Physical Chemistry, 1995. The "Distance Scaling Method" es un procedimiento para deformar la hipersuperficie de energía potencial. El procedimiento es una extensión del "procedimiento de ecuación de difusión" y adolece de los mismos inconvenientes. Este no es un procedimiento de búsqueda determinístico y en consecuencia no asegura la mejor estructura cada vez. También es muy computacionalmente intensivo y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing*" de J. Lee y H. A. Scheraga, Biopolymers, 1998 combina los aspectos esenciales del procedimiento de construcción y un algoritmo genético. Busca la totalidad del espacio conformacional en sus etapas iniciales y luego estrecha la búsqueda a regiones más pequeñas con baja energía. Sin embargo, la búsqueda de la totalidad del espacio conformacional requiere gran cantidad de tiempo de computación. En consecuencia es muy intensivo computacionalmente y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*Protein structure can be predicted by global optimization of a potential energy function*" de A. Liwo, J. Lee, D. R Ripoll, J. Pillardy y H. A. Scheraga, Proceedings of the National Academy of Sciences, 1999 usa un abordaje

jerárquico basado en dos desarrollos recientes, un campo de fuerza de residuo unido (UNRES) y el procedimiento CSA. Se realiza una extensa búsqueda conformacional con CSA usando el campo de fuerza UNRES. Este adolece de los mismos inconvenientes que el "procedimiento de empalme de espacio conformacional". En otras palabras, es muy computacionalmente intensivo y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*Taboo Search: An approach to the Multiple Minima Problem*", de D. Cvijiovic y J. Klinowski, Science, 1995 describe un procedimiento, Taboo Search (TS), sobre la base de la búsqueda de tabú de Glover para funciones discretas para solucionar el problema de mínimos múltiples para funciones continuas. El procedimiento no indica ninguna aplicación a estructuras moleculares. También es muy computacionalmente intensivo y en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

"*TRUST: A deterministic algorithm for Global Optimization*" de J. Barthen, V. Protopopescu y D. Reister, Science, 1997 sugiere un abordaje, el algoritmo *Terminal Repeller Unconstrained Sub-energy Tunneling* (TRUST) para solucionar los problemas continuos de optimización global. El procedimiento sin embargo no indica ninguna aplicación a estructuras moleculares y también es muy computacionalmente intensivo. En consecuencia el procedimiento en consecuencia muy oneroso en términos de tiempo de computación. Tampoco hay ninguna mención sobre el uso de MOLS en cualquier forma del procedimiento.

Objetos de la invención

El principal objetivo de la presente invención consiste en proveer un procedimiento para construir óptimos modelos tridimensionales de estructuras moleculares a partir del conocimiento de sus estructuras químicas, lo cual obvia los inconvenientes antes mencionados, tal como se detalla a continuación.

Otro objeto de la presente invención consiste en proveer un procedimiento y un sistema para construir estructuras de energía mínima de macromoléculas biológicas tales como proteínas y ácidos nucleicos.

Otro objeto de la presente invención consiste en aplicar el procedimiento de la invención para hallar el mejor valor de una función.

Otro objeto de la presente invención consiste en aplicar el procedimiento de la invención para hallar el mejor valor de una función de energía potencial de una molécula.

Otro objeto de la presente invención consiste en hacer que el procedimiento de determinación de estructura sea simple y directo y de bajo costo de uso, en términos de tiempo de computación.

Otro objeto de la presente invención consiste en aplicar el procedimiento de la invención a una amplia variedad de problemas en los cuales se define un sistema físico o químico definido como función de un conjunto de parámetros y la invención se usa para hallar la mejor configuración del sistema a través de la búsqueda del mejor conjunto de valores de los parámetros que dan el mejor valor de la función.

Otro objeto de la presente invención consiste en proveer un procedimiento que se pueda confiar fácilmente con otros procedimientos conocidos tales como búsqueda de gradientes, empalme simulado, etc.

Otro objeto de la presente invención consiste en hacerlo fácilmente programable para correr en computadoras paralelas, por lo que se ahorra tiempo valioso.

Sumario de la invención

En consecuencia, la presente invención, provee un procedimiento para determinar el menor valor de la función de energía potencial conformacional de una molécula útil para construir modelos de estructuras moleculares tridimensionales usando un nuevo algoritmo que define la mejor conformación como la correspondiente al menor valor para la función de energía potencial conformacional, en donde dicho procedimiento comprende las etapas de:

a) proveer la representación de una molécula en términos de las coordenadas de los átomos que comprenden la molécula;

b) seleccionar un conjunto de conformaciones elegidas de todas las posibles conformaciones usando un ordenamiento combinatorio de Cuadrados Latinos Mutuamente Ortogonales (MOLS) de orden N, en donde dichos MOLS se construyen sobre la base de parámetros conformacionales en la molécula, en donde hay N^2 subcuadrados en cada conjunto de MOLS, en donde N es igual a n o m, el que sea mayor, en donde n es la cantidad de parámetros que definen la conformación de la molécula y m es la cantidad de valores de muestra a lo largo de cada parámetro conformacional y el tamaño de etapa es R/m, en donde R es el rango del parámetro, y cada subcuadrado corresponde a una posible conformación de dicha molécula;

c) llenar los subcuadrados de los MOLS con conformaciones de dicha molécula, con una conformación en cada subcuadrado, en donde la conformación en cualquier subcuadrado dado está representado por los valores de los parámetros conformacionales en dicho subcuadrado;

d) elegir los valores de los parámetros conformacionales en los subcuadrados al especificar primero la cantidad de parámetros, n, y el rango de cada uno de los n parámetros, y el tamaño de etapa para la cual se muestrean los valores de cada parámetro, en donde hay un total de m etapas a lo largo de cada parámetro, de manera tal que se especifican $\theta_{r,s}$, r = 1 a n y s = 1 a m como el conjunto de valores de ingreso para los parámetros conformacionales;

e) rotular los subcuadrados de dichos MOLS mediante los índices (u,t), en donde u y t son números enteros, u = 1 a N y t = 1 a N, y el valor del r-ésimo ángulo de un determinado subcuadrado (u,t) dado se elige del conjunto de valores de ingreso $\theta_{r,s}$ como $\theta^{r}_{u,t} = \theta_{r,s}$, en los cuales el índice u está dado para u = {(t-1) (r-1) + (s-1)} modulo (N);

f) repetir el procedimiento de llenar los subcuadrados con los valores de los parámetros elegidos para todos los valores de r = 1 a n, s = 1 a m, y t = 1 a N, en donde u se calcula cada vez mediante la fórmula dada en la etapa (e), por la cual al final de cada procedimiento, en donde cada subcuadrado de los N^2 subcuadrados de dicho MOLS es llenado con un conjunto de n valores, en donde cada valor corresponde a uno de los n parámetros conformacionales, en con secuencia el conjunto de n valores en cada subcuadrado define una conformación de dicha molécula, en donde hay N^2 conformaciones;

g) calcular la función de energía potencial mediante procedimientos conocidos para cada una de las N^2 conformaciones de dicha molécula tal como se representa en los N^2 subcuadrados de dicho conjunto de MOLS, en donde dichas funciones están representadas por

$$V_{u,t} = V(\theta^{r}_{u,t})$$

h) analizar el conjunto de N^2 valores de la función de energía potencial a fin de determinar el mejor valor de cada parámetro conformacional;

i) analizar por determinación del valor promedio de la función de energía potencial a cada valor de cada parámetro, al tomar el promedio de los valores de la función de energía potencial para todas las N conformaciones en las cuales se usa el valor del parámetro conformacional para definir las conformaciones, por lo que hay n x m valores promedio, en donde los valores promedio están dados por

$$\langle V \rangle^{r}_{t} = \frac{\sum_{1} V(\theta^{r}_{u,t}) \exp(-V(\theta^{r}_{u,t})/kT)}{\sum \exp(-V(\theta^{r}_{u,t})/kT)}$$

j) escanear el conjunto de n x m valores promedio de la función de energía potencial con el fin de obtener, para cada parámetro conformacional, el valor de dicho parámetro que corresponde al menor valor promedio de la función de energía potencial, el cual es seguido del uso de conjuntos de n valores de los n parámetros conformacionales, un valor para cada parámetro, a fin de definir la mejor conformación de dicha molécula que tiene el menor valor de función de energía potencial.

k) confirmar los resultados o ubicar otra estructura óptima, de ser requerido, mediante la elección de una forma diferente de disponer los valores de los parámetros conformacionales y repetir las etapas e a k.

En una forma de realización de la invención, el procedimiento comprende obtener los óptimos de varias funciones no lineales simples, cada una con un único óptimo conocido en el rango del parámetro.

En otra forma de realización de la invención, el procedimiento se lleva a cabo mediante la búsqueda del mejor valor de diferentes tipos de funciones de muchos parámetros.

En otra forma de realización de la invención, el procedimiento se lleva a cabo mediante la obtención de la mejor estructura de moléculas pequeñas tales como péptidos, nucleótidos, etc.

En otra forma de realización de la invención, el procedimiento se lleva a cabo mediante la obtención de la mejor estructura de macromoléculas biológicas tales como proteínas y ácidos nucleicos, etc.

En otra forma de realización de la invención, el procedimiento se corre en paralelo en diversas computadoras simultáneamente para así reducir aún más el tiempo de computación.

Breve descripción de los dibujos adjuntos

La Figura 1 es un ejemplo de un conjunto de Cuadrados Latinos Mutuamente Ortogonales (MOLS) a fin de explicar mejor el procedimiento. El diagrama muestra un ejemplo de tres MOLS de orden 7, es decir n = 3, m = 7, N = 7. Símbolos en el primer Cuadrado Latino: a1, a2, a3, a4, a5, a6, a7. Cada uno de estos se repite 7 veces para dar un total de 49 símbolos, los cuales se han dispuesto en un LS. Símbolos en el segundo LS: b1, b2, b3, b4, b5, b6, b7. El segundo LS es ortogonal respecto del primero. Nótese que cada apareamiento de un símbolo proveniente del primer Cuadrado con uno del segundo ocurre exactamente una vez. Símbolos en el tercer LS: c1, c2, c3, c4, c5, c6, c7. Este es ortogonal a ambos otros Cuadrados. Por claridad en esta figura 3 diferentes conjuntos de N símbolos. Se puede usar el mismo conjunto de N símbolos y construir N-1 MOLS de orden N. Se ha destacado uno de los subcuadrados del conjunto de MOLS, sus símbolos son a7 del primer LS, b1 del segundo LS y c5 del tercer LS. En la presente aplicación, cada uno de dichos subcuadrados representa una posible conformación de la molécula. El procedimiento MOLS requiere evaluar la función potencial en todos estos N^2 puntos en el espacio de conformación.

La Figura 2a provee la estructura de la molécula pentapéptido [Met5] encefalina (Tyr-Gly-Gly-Phe-Met) tal como se obtiene mediante una aplicación del presente procedimiento MOLS tal como se describe en el ejemplo 1. El valor de la función de energía potencial para esta estructura es -13,61 kcal/mol.

La Figura 2b muestra por comparación la estructura de la molécula tal como es informado por Z. Li y H. A. Scheraga en *Proceedings of the National Academy of Sciences* (USA), Volumen 84, páginas 6611-6615. Es obvio para la comparación que las dos estructuras son iguales y que el presente procedimiento MOLS identifica con éxito la correcta estructura de la molécula. El presente procedimiento MOLS identifica esta estructura y 99 otras estructuras en un tiempo total de computación de 65 segundos en un procesador Pentium III que opera a 650 MHz. Este tiempo es inferior al tiempo requerido por el procedimiento de Li y Scheraga.

La Figura 3a muestra otra estructura de la molécula pentapéptido [Met5] encefalina (Tyr-Gly-Gly-Phe-Met) tal como se obtiene por aplicación del presente procedimiento MOLS tal como se describe en el ejemplo 1. El valor de la función de energía potencial para esta estructura es -8,53 kcal/mol.

La Figura 3b muestra por comparación la estructura determinada experimentalmente de la molécula informada por Griffin, *et al*., en *Proceedings of the National Academy of Sciences* (CSA), volumen 83, páginas 3272-3276. Es obvio para la comparación que las dos estructuras son iguales y que el presente procedimiento MOLS identifica con éxito la correcta estructura de la molécula. El presente procedimiento MOLS identifica esta estructura y 99 otras estructuras, incluso la mostrada en el dibujo 2 anterior, en un tiempo total de computación de 65 segundos en un procesador Pentium III que opera a 650 MHz.

La Figura 4 muestra el gráfico de Ramachandra para cien estructuras diferentes, incluso las estructuras mostradas en la Figura 2 y en la Figura 3 anteriores, obtenidas al repetir el procedimiento con un conjunto diferente de valores de entrada cada vez. El dibujo muestra claramente que todas las estructuras son conformacionalmente las mejores estructuras estudiadas mediante el mapa de Ramachandra, dado que sus conformaciones corresponden todas a las regiones permitidas del mapa de Ramachandra.

Las Figuras 5a y 5b muestran dos estructuras de energía mínima del pentapéptido $(Aib)_5$ tal como se obtiene por el presente procedimiento MOLS (en verde). Los valores de la función de energía potencial para estas estructuras son 5,19 kcal/mol y 9,41 kcal/mol respectivamente. Por comparación, estas dos estructuras se superponen a las estructuras (en rojo) de la misma molécula obtenida experimentalmente por Benedetti *et al*. e informados en *Journal of the Chemical Society*, volumen 104, páginas 2437-2444, y por Shamala *et al*. e reportados en *Journal of Chemical Society, Chemical Communications*, año 1978 página 996. La comparación demuestra que el presente procedimiento MOLS produce estructuras correctas. Estas dos estructuras, junto con otras 998 fueron identificadas en un tiempo total inferior a 1 hora mediante el uso del procesador Pentium III a 650 MHz.

La Figura 6 demuestra que el gráfico de Ramachandra para mil estructuras diferentes, incluso las estructuras mostradas en la Figura 5 anterior, obtenidas por el presente procedimiento MOLS, al repetir el procedimiento con un conjunto diferente de valores de entrada cada vez. El dibujo muestra claramente que todas las estructuras son conformacionalmente las mejores estructuras estudiadas mediante el mapa de Ramachandra, dado que todas sus conformaciones correspondan a las regiones permitidas del mapa de Ramachandra. Estas 1000 estructuras fueron obtenidas en un tiempo total de menos de 60 minutos en un procesador Pentium III a 650 MHz.

La Figura 7 provee un diagrama de flujo para explicar el procedimiento.

Descripción detallada de la invención

El procedimiento de la invención se explica mejor en términos de un conjunto de etapas. La construcción del conjunto de MOLS comprende un Cuadrado Latino (LS) de orden N definido como un conjunto de N símbolos, dispuestos en un cuadrado de N x N, de manera tal que cada símbolo aparece exactamente una vez en cada hilera y una vez en cada columna. Dos LS son ortogonales si, cuando se superponen, cada símbolo del primer cuadrado aparece una y solamente una vez, con cada símbolo del segundo cuadrado. Un conjunto de MOLS es un conjunto de LS, cada par de los cuales es ortogonal (Figura 1). Se ha demostrado en la técnica previa que si N es una potencia

prima, se pueden construir N-1 MOLS de orden N, como sigue. Los N símbolos (pertenecientes a un campo de Galois) se escriben como b_i, en donde i = 1 a N. Si es el elemento (i, j)-ésimo del LS rotulado con e, entonces

$$a_{ij}^{e} = b_e x\, b_i + b_j \qquad (1)$$

para i = 1 a N; j = 1 a N; e = 1, N-1. En el presente algoritmo se hace una correspondencia entre los símbolos b y los parámetros conformacionales de la molécula. Por ejemplo, pueden ser los ángulos de torsión θ_i, para i = 1 a n, que caracterizan la conformación de un oligopéptido. La mejor conformación de la molécula se define como la que tiene el mínimo valor de la función de energía potencial, la cual es una función de los parámetros conformacionales. En otras palabras, si se encuentra el conjunto de parámetros conformacionales que generan un valor mínimo para la función de energía potencial, entonces dicho conjunto de parámetros conformacionales corresponderá a la mejor conformación de la molécula. Cada parámetro conformacional se especifica como capaz de adoptar m valores diferentes en un rango particular. Si r es el índice para estos parámetros conformacionales y s es el índice para los valores adoptados por cada parámetro, entonces $\theta_{r,s}$, r = 1 a n; s = 1 a m son los conjuntos de valores de entrada, y la búsqueda para la conformación se debe llevar a cabo entre las nm combinaciones de estos valores. En consecuencia, por ejemplo, $\theta_{1,1}$ podría referirse al primer valor del primer parámetro, $\theta_{1,2}$ al segundo valor del primer parámetro, $\theta_{2,1}$ al primer valor del segundo parámetro, $\theta_{2,2}$ al segundo valor del segundo parámetro, y así sucesivamente hasta $\theta_{n,m}$, el cual se referiría al último valor del último parámetro. Obviamente habría un conjunto de m x n valores de entrada. Para construir el conjunto de MOLS, cada parámetro conformacional se fija para corresponder a un LS, y se aplica la ecuación (1) a los índices de los parámetros conformacionales. Se considera que los índices r y s forman n conjuntos de Galois, cada uno con m elementos, es decir el par de índices (r,s) formará el s-ésimo elemento en el r-ésimo conjunto. Los subcuadrados de los MOLS se rotulan mediante el par de índices (u,t), en donde ambos u y t adoptan valores enteros desde 1 hasta N, el valor del ángulo de torsión r-ésimo en el subcuadrado (u,t) se puede elegir del conjunto de valores de entrada $\theta_{r,s}$

$$\theta^{r}_{u,t} = \theta_{r,s} \qquad (2)$$

en donde el índice u está dado por

$$u = \{(t-1)(r-1) + (s-1)\}\text{modulo}(N)$$

El valor de N se elige de manera tal que (i) N-1 ≥ máximo (m,n) y (ii) N es una potencia prima. Esta última condición requiere que el la práctica el tamaño de la etapa en la grilla de muestreo se elija como R/N, en donde R es el rango específico para los parámetros conformacionales. La aplicación de este procedimiento para todos los valores de r = 1 a n; s = 1 a N y t = 1 a N da por resultado un conjunto de n MOLS de orden N. Estos MOLS, cuando se superponen, darán N^2 subcuadrados, en donde cada subcuadrado contiene un conjunto de valores para los n parámetros conformacionales (Figura 1). Cada conjunto de valores de los parámetros, rotulados por el par de índices (u,t), representa una conformación posible de la cadena de polipéptido y, en consecuencia, un punto en el espacio conformacional de n dimensiones. El procedimiento de la presente invención demuestra que el muestreo de la superficie potencial de la molécula en los N^2 puntos especificados por los subcuadrados, de los posibles nm, es suficiente para construir un mapa de todo el espacio, y que este mapa luego puede ser usado para realizar una rápida búsqueda del óptimo. La hipótesis es avalada por el hecho de que el conjunto de MOLS contiene implícitamente la totalidad de las proyecciones bidimensionales posibles del espacio n-dimensional. La siguiente etapa consiste en el muestreo de la hipersuperficie de energía en el espacio conformacional, el cual se lleva a cabo mediante el cálculo de la energía potencial

$$V_{u,t} = V(\theta^{r}_{u,t})$$

en cada uno de los N^2 puntos u = 1 a N; t = 1 a N. Después de completar esta etapa y de calcular los N^2 valores de la función de energía potencial, la siguiente tarea consiste en recuperar el mapa energético del espacio conformacional. Para lograrlo, se construyen n representaciones de una dimensión de la variación de la función de energía potencial V a lo largo de cada uno de los parámetros conformacionales. El efecto de fijar un parámetro conformacional particular en un valor particular se estima al tomar el promedio de V para los N puntos en los cuales el valor de dicho parámetro aparece en el MOLS, es decir

$$\langle V \rangle^{r}_{s} = \frac{\sum_{t} V(\theta^{r}_{u,t}) \exp(-V(\theta^{r}_{u,t})/kT)}{\sum \exp(-V(\theta^{r}_{u,t})/kT)}$$

en donde k es la constante de Boltzmann y T es la temperatura. Una vez más, se usa la ecuación (2) para calcular el valor de u para cada valor de r, s y t. Dado que la cantidad de etapas m a lo largo de cada parámetro se fijan como iguales a N, habrá N de dichos valores para cada uno de los r = 1 a n parámetros conformacionales. Los N

valores promedio para determinado parámetro conforman una representación del comportamiento del V potencial como función de dicho parámetro. La etapa final es una inspección de cada representación unidimensional, lo cual en consecuencia revelará el valor óptimo para la respectiva torsión. En consecuencia, si

$$V^{r}_{s=\omega_r} = \underset{\text{para } s}{\text{óptimo}} \; ((V)^{r}_{s}) \qquad (3)$$

entonces $\theta_{r,\omega r}$ es el valor óptimo para el parámetro r. El conjunto de valores óptimos para $\theta_{r,\omega}$, r = 1 a n definirá entonces la conformación óptima de la molécula. Esto completa el procedimiento.

Los índices r y s usados en la construcción de MOLS se pueden asignar a los m valores de los n parámetros en $(m!)^n$ diferentes formas. De acuerdo con la invención, sin importar la forma en la cual se realiza la asignación, el procedimiento dará por resultado una estructura óptima. En consecuencia, mediante la elección de diferentes asignaciones de los ángulos, se pueden repetir los cálculos a fin de verificar la constancia de los resultados obtenidos.

Novedad respecto a la técnica previa

1. Los Cuadrados Latinos Mutuamente Ortogonales (MOLS) no se han usado previamente para hallar la mejor estructura molecular o biomolecular.

2. Los MOLS no se han usado previamente para hallar el mejor valor de una función.

3. Los MOLS no se han usado previamente para hallar el mejor valor global de una función.

4. La ecuación 2 no se ha establecido o usado previamente.

5. El procedimiento de tomar promedios tal como se da en la ecuación (3) no se ha usado previamente.

Los siguientes ejemplos se dan a modo de ilustración de la presente invención y no se deben interpretar como limitantes del alcance de la presente invención.

Ejemplo 1

Como ejemplo, se aplica el procedimiento de la invención al pentapéptido [Met5] encefalina (Tyr-Gly-Gly-Phe-Met). Ignorando los ángulos de torsión de las cadenas laterales, que se mantuvieron fijos en los valores hallados en la bibliografía, hay 8 ángulos de torsión del esqueleto. Estos se consideraron los parámetros conformacionales, con rango entre 0 y 360º. La obtención de cada uno de ellos a intervalos de 10º requeriría 36 puntos a lo largo de cada dimensión. El espacio total de búsqueda es en consecuencia de 36^8 o aproximadamente 2,8 x 10^{12} puntos y una búsqueda exhaustiva para la estructura óptima implicaría el cálculo de la energía en cada uno de dichos puntos para elegir el mejor. En el presente procedimiento, se calcula la energía en sólo 36^2 puntos. Sin embargo, dado que el orden de MOLS siempre es un número primo y dado que 36 no es un número primo, en realidad se buscan 37^2 puntos. En la primera etapa se especifica el conjunto de valores para cada ángulo de torsión que se ha de considerar. En consecuencia, los valores para el primer ángulo de torsión pueden ser 0º, 10º, 20º, 30º, 360º. De modo similar, el segundo ángulo de torsión se debe considerar en los valores 0º, 10º, 360º. De modo similar, también se especifican los puntos de muestreo de los demás ángulos de torsión. En consecuencia, se especifican los valores $\theta_{r,s}$ para r = 1,8; s = 1,37. Ahora se usa la ecuación (2) para distribuir 37 copias de estos 37 valores de cada uno de los 8 ángulos de torsión entre los 37^2 subcuadrados a fin de conformar un conjunto de 8 Cuadrados Latinos Mutuamente Ortogonales de orden 8. Los 8 valores en cada subcuadrado luego forman una conformación posible de la molécula. Nótese que algunas de estas conformaciones pueden ser conformaciones “imposibles”, en el sentido que los átomos pueden interpenetrarse. Sin embargo, la presente invención permite derivar la conformación óptima a partir de este conjunto de 37^2 conformaciones. La siguiente etapa en el procedimiento consiste en calcular la energía potencial semiempírica en cada uno de estos 37^2 puntos. Una vez calculadas estas energías, la siguiente etapa comprende verificar la energía promedio para cada valor de cada ángulo de torsión. Por ejemplo, para el primer valor del primer ángulo de torsión es 0º. Hay 37 conformaciones de la molécula para las cuales el valor del primer ángulo de torsión se ha fijado en 0º y los valores de la función de energía potencial de todas estas conformaciones se han determinado en el procedimiento anterior. Se toma el promedio ponderado de Boltzmann de estos 37 valores de energía, y en consecuencia se obtiene un único número que especifica el efecto de fijar dicho ángulo respecto de dicho valor, es decir el efecto de fijar la primera torsión en el valor 0º. De modo similar, se obtiene la energía promedio cuando se fija la primera torsión en 10º, 20º, etc. Es importante destacar que la propiedad de MOLS para cada situación se repite 37 veces (o N veces, en donde N es el orden de MOLS). Dado que hay 37 situaciones para la torsión, habrá 37 energías promedio. La siguiente etapa comprende inspeccionar estos 37 promedios e identificar una situación de esta torsión que ha producido la menor energía promedio. Este será el valor óptimo para este ángulo de torsión. El procedimiento se repite para el siguiente ángulo de torsión y también se determina su valor óptimo de manera exactamente similar al primero. Nótese que nuevamente, debido a la propiedad de MOSL, cada valor de la segunda torsión también se repite 37 veces. Todo el procedimiento se vuelve a repetir para los demás ángulos de torsión. Al final de esta etapa, la invención ha determinado individualmente la mejor situación para cada uno de los 8 ángulos de torsión. Cuando ahora se construye la molécula con estos 8 valores de ángulo de

torsión identificados mediante la invención, esta será la conformación óptima de [Met^5] encefalina. La totalidad del procedimiento se repite con la elección de $\theta_{r,s}$ de diferente manera. Ahora en lugar de especificar $\theta_{1,1}$ = 0,0, $\theta_{1,2}$ = 10,0 etc., los valores especificados para $\theta_{1,1}$ = 10,0, $\theta_{1,2}$ = 0,0 etc. La asignación se realiza usando un generador de números al azar. La etapa se repite para dar otra conformación óptima para la molécula. Esto se puede repetir varias veces para dar tantas estructuras óptimas como sean requeridas. En este ejemplo, se generaron cien estructuras óptimas. Cada estructura era el resultado de la aplicación del procedimiento anterior a un conjunto de entrada diferente de ángulos de torsión $\theta_{r,s}$, los cuales se obtuvieron mediante una asignación diferente de los índices a los valores de ángulos de torsión. Las asignaciones se eligieron mediante el uso de un generador de números al azar. Las energías correspondientes a todas estas estructuras son bajas, en donde la mayor es de 72,3 kcal/mol y la inferior es de -14,0 kcal/mol, con un promedio de -7,34 kcal/mol para la totalidad de 100 estructuras. La mejor estructura se muestra en la Figura 2 y la Figura 3 y el gráfico conformacional de Ramachandra para la totalidad de las 100 estructuras se muestra en la Figura 4.

Ejemplo 2

Como otro ejemplo de la invención, para mostrar su aplicación para funciones arbitrarias y no sólo para funciones de energía potencial basadas en parámetros conformacionales, el procedimiento se aplica a una función artificial con un mínimo conocido como sigue:

$$F = 2^{(a-5)} \times 3^{|b-2|} \times 4^{|c-9|} 5^{|d-6|} \times 6^{|e-8|} \times 6^{(f-8)} \times 5^{(g-6)} \times 4^{(h-9)} \times 3^{(i-2)} \times 2^{|j-5|} \times (k-7.1)^4$$

Esta es una función de las 11 variables a, b, c, d, e, f, g, h, i, j, k, todas ellas muestras tomadas en el rango de 1,0 a 13,0, en etapas de 1,0. El mínimo de la función en este espacio está en a = 1,0, b = 2,0, c = 9,0, d = 6,0, e = 8,0, f = 1,0, g = 1,0, h = 1,0, i = 1,0, j = 5,0, k = 7,0. Se construye un conjunto de 11 MOLS de orden 13 usando los valores 1,0, 2,0, ..., 13,0 como rango de valores para cada unas de las 11 variables. La función se evalúa en cada uno de los puntos especificados por los 13 x 13 subcuadrados. Se toman los promedios y el mínimo se identifica como 1,0, 1,0, 8,0, 5,0, 8,0, 2,0, 1,0, 1,0, 1,0, 5,0, 7,0. Este resultado es muy cercano al mínimo teórico dado con anterioridad.

Ejemplo 3

En este ejemplo, la molécula elegida comprende el pentapéptido $(Aib)_5$. Como en el ejemplo 1, la cantidad de ángulos de torsión libres en esta molécula es 8. Nuevamente se eligen estos como los parámetros de conformacional, con el rango fijado desde 0 hasta 360º. En las etapas de 10º hay 37 puntos a lo largo de cada dimensión y el orden de los MOLS se elige en 37. Los 37^2 puntos conformacionales se especifican usando MOLS, se calculan funciones de energía potencial, se toman los promedios y se identifica la conformación óptima tal como se describe en el ejemplo 1. Esto se repite 1000 veces en un tiempo de computadora total de aproximadamente 1hora para todas las 1000 conformaciones. La totalidad de ellas son buenas conformaciones consideradas por sus bajos valores de energía. Algunas de ellas corresponden a estructuras bien conocidas tales como hélices alfa o bucles. La Figura 5 muestra las dos mejores estructuras y la Figura 6 el gráfico conformacional de Ramachandra de las 1000 estructuras.

Síntesis de observaciones: La invención opera muy bien para diferentes tipos de funciones. Particularmente en análisis de conformación molecular, la invención permite obtener el mapa de la totalidad del espacio conformacional en un tiempo razonable, a fin de identificar estructuras físicamente razonables.

Las principales ventajas de la presente invención son

1. La principal ventaja de la presente invención consiste en que provee un procedimiento para construir modelos tridimensionales óptimos de estructuras moleculares a partir del conocimiento de sus estructuras químicas, lo cual obvia el inconveniente de los procedimientos previos.

2. Otra ventaja de la presente invención consiste en que provee un procedimiento y un sistema para construir estructuras de mínima energía de macromoléculas biológicas tales como proteínas y ácidos nucleicos.

3. Otra ventaja de la presente invención consiste en que se puede aplicar para hallar el mejor valor de una función.

4. Otra ventaja de la presente invención consiste en que se puede aplicar para hallar el mejor valor de una función de energía potencial de una molécula.

5. Otra ventaja de la presente invención consiste en que es simple y directa y de bajo costo de uso, en términos de tiempo de computación.

6. Otra ventaja de la presente invención consiste en que se puede aplicar a una amplia variedad de problemas en los cuales se define un sistema físico o químico como función de un conjunto de parámetros y la invención se usa para hallar la mejor configuración del sistema a través del hallazgo del mejor conjunto de valores de los parámetros que dan el mejor valor de la función.

7. Otra ventaja de la presente invención consiste en que se puede combinar con otros procedimientos conocidos tales como búsqueda de gradiente, empalme simulado, etc.

8. Otra ventaja de la presente invención consiste en que se puede programar fácilmente para correr en computadoras paralelas, lo cual ahorra tiempo valioso.

REIVINDICACIONES

1. Un procedimiento implementado por computadora para determinar el menor valor de la función de energía potencial conformacional de una molécula útil para construir modelos de estructuras moleculares tridimensionales usando un nuevo algoritmo que define la mejor conformación como aquella correspondiente al menor valor de la función de energía potencial conformacional, en donde dicho procedimiento comprende las etapas de:

(a) proveer la representación de una molécula en términos de las coordenadas de los átomos que comprenden la molécula, por procedimientos conocidos;

(b) seleccionar un conjunto de conformaciones elegidas entre todas las conformaciones posibles mediante el uso de una disposición combinatoria de n Cuadrados Latinos Mutuamente Ortogonales (MOLS) de orden N, en donde dichos MOLS se construyen sobre la base de los parámetros conformacionales de la molécula, en donde hay N^2 subcuadrados en cada conjunto de MOLS, en donde N es igual a n o m, el que sea mayor, en donde n es la cantidad de parámetros que define la conformación de la molécula y m es la cantidad de valores tomados como muestra a lo largo de cada parámetro conformacional y en donde el tamaño de la etapa es R/m, en donde R está en el rango del parámetro, y cada subcuadrado corresponde a una posible conformación de dicha molécula;

(c) llenar los subcuadrados de MOLS con conformaciones de dicha molécula, con una conformación en cada subcuadrado, en donde la conformación en cualquier subcuadrado dado está representado por los valores de los parámetros conformacionales en dicho subcuadrado;

(d) elegir los valores de los parámetros conformacionales en los subcuadrados al especificar primero la cantidad de parámetros, n, y el rango de cada uno de los n parámetros, y el tamaño de las etapas a las cuales se toman muestras de los valores de cada parámetro, en donde hay un total de m etapas a lo largo de cada parámetro, de manera tal que se especifica que $\theta_{r,s}$, r = 1 a n y s = 1 a m son el conjunto de valores de entrada para los parámetros conformacionales;

(e) rotular los subcuadrados de dichos MOLS mediante los índices (u, t), en donde u y t son números enteros, u = 1 a N y t = 1 a N, y el valor del r-ésimo ángulo de torsión en un subcuadrado (u, t) dado se elige del conjunto de valores de entrada $\theta_{r,s}$, con $\theta^{r}{}_{u,t}$, = $\theta_{r,s}$, en donde el índice u está dado por u = {(t-1) (r-1) + (s-1)} módulo (N);

(f) repetir el procedimiento de llenar los subcuadrados con los valores de los parámetros elegidos de todos los valores de r = 1 a n, s = 1 a m, y t = 1 a N, en donde u se calcula cada vez mediante la fórmula dada en la etapa (e), por lo cual al final del procedimiento, cada subcuadrado de los N^2 subcuadrados de dicho MOLS se llenan con un conjunto de n valores, en donde cada valor corresponde a uno de los n parámetros conformacionales, por lo que el conjunto de n valores en cada subcuadrado define una conformación de dicha molécula, por lo que hay N^2 conformaciones;

(g) calcular la función de energía potencial mediante procedimientos conocidos para cada una de las N^2 conformaciones de dicha molécula tal como se representa en los N^2 subcuadrados de dicho conjunto de MOLS, en donde dichas funciones están representadas por

$$V_{u,t} = V(\theta^{r}_{u,t})$$

(h) analizar el conjunto de N^2 valores de la función de energía potencial a fin de determinar el mejor valor de cada parámetro conformacional;

(i) analizar por determinación del valor promedio de la función de energía potencial para cada valor de cada parámetro, al tomar el promedio de los valores de la función de energía potencial en todas las N conformaciones en las cuales se usa dicho valor del parámetro conformacional para definir las conformaciones, en donde en consecuencia hay n x m valores promedio, en donde los valores promedio están dados por

$$\langle V \rangle^{r}_{s} = \frac{\sum_{t} V(\theta^{r}_{u,t}) \exp(-V(\theta^{r}_{u,t})/kT)}{\sum_{t} \exp(-V(\theta^{r}_{u,t})/kT)}$$

(j) escanear el conjunto de n x m valores promedio de la función de energía potencial a fin de obtener, para cada parámetro conformacional, el valor de dicho parámetro que corresponde al menor valor promedio de la función de energía potencial, el cual es seguido mediante el uso de un conjunto de n valores de los n parámetros conformacionales, un valor para cada parámetro, a fin de definir la mejor conformación de dicha molécula que tiene el menor valor de la función de energía potencial; y

(k) confirmar los resultados o localizar otra estructura óptima, si se requiere, mediante la elección de una forma diferente de ordenar los valores de los parámetros conformacionales y repetir las etapas (e) a (k).

2. Un procedimiento implementado por computadora de acuerdo con la reivindicación 1, en el que dicho procedimiento comprende obtener la óptima de varias funciones no lineales simples, cada una con un único óptimo conocido en el rango del parámetro.

3. Un procedimiento implementado por computadora de acuerdo con la reivindicación 1, en el que dicho procedimiento se lleva a cabo mediante la búsqueda del mejor valor de diferentes tipos de funciones de muchos parámetros.

4. Un procedimiento implementado por computadora de acuerdo con la reivindicación 1, en el que la molécula comprende una molécula pequeña seleccionada en que péptidos y nucleótidos.

5. Un procedimiento implementado por computadora de acuerdo con la reivindicación 1, en el que la molécula comprende una macromolécula biológica seleccionada entre proteínas y ácidos nucleicos.

6. Un procedimiento implementado por computadora de acuerdo con la reivindicación 1, que se ejecuta en paralelo en varias computadoras simultáneamente para reducir aún más el tiempo de computación.

Figura 1

a1 b1 c1	a2 b2 c2	a3 b3 c3	a4 b4 c4	a5 b5 c5	a6 b6 c6	a7 b7 c7
a2 b7 c6	a3 b1 c7	a4 b2 c1	a5 b3 c2	a6 b4 c3	a7 b5 c4	a1 b6 c5
a3 b6 c4	a4 b7 c5	a5 b1 c6	a6 b2 c7	a7 b3 c1	a1 b4 c2	a2 b5 c3
a4 b5 c2	a5 b6 c3	a6 b7 c4	a7 b1 c5	a1 b2 c6	a2 b3 c7	a3 b4 c1
a5 b4 c7	a6 b5 c1	a7 b6 c2	a1 b7 c3	a2 b1 c4	a3 b2 c5	a4 b3 c6
a6 b3 c5	a7 b4 c6	a1 b5 c7	a2 b6 c1	a3 b7 c2	a4 b1 c3	a5 b2 c4
a7 b2 c3	a1 b3 c4	a2 b4 c5	a3 b5 c6	a4 b6 c7	a5 b7 c1	a6 b1 c2

Figura 2 a

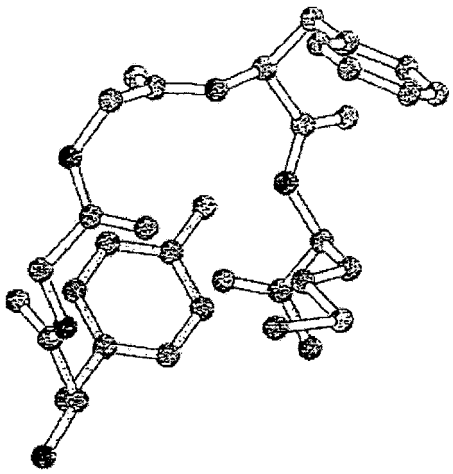


Figura 2 b

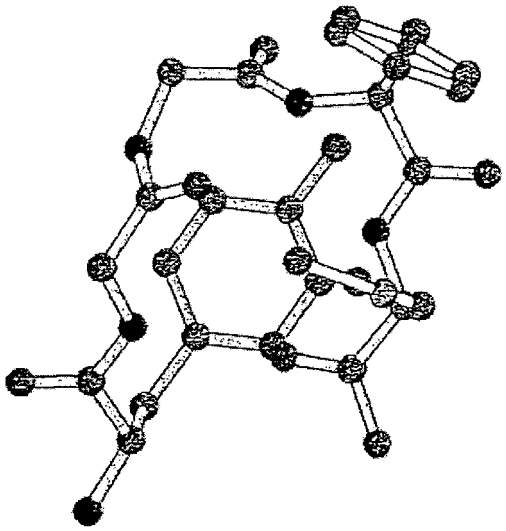


Figura 3 a

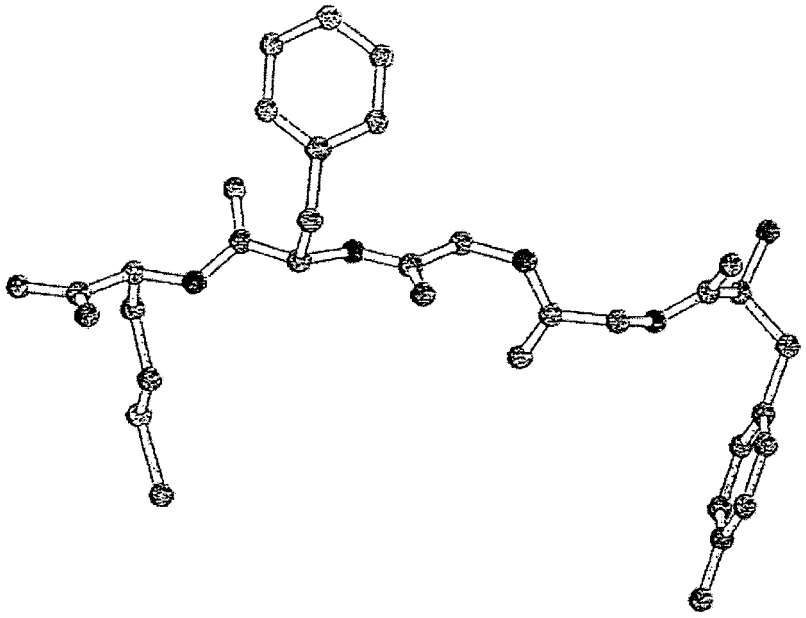


Figura 3 b

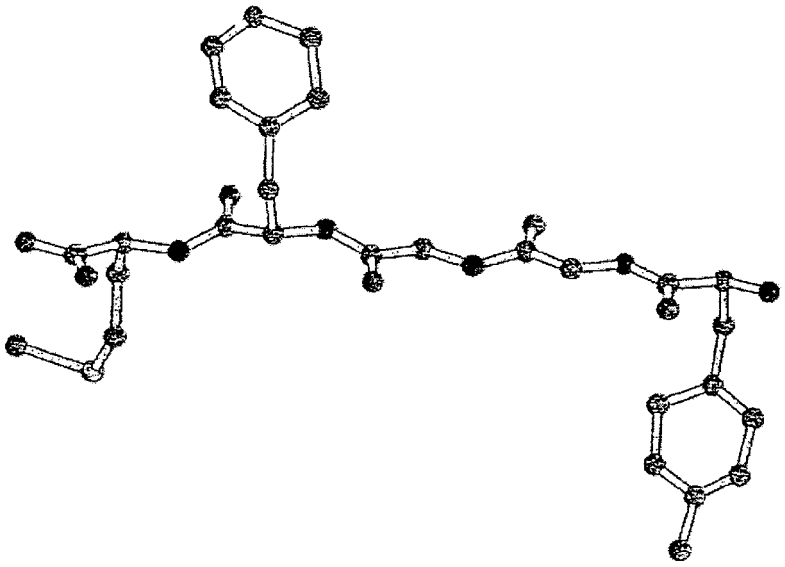


Figura 4

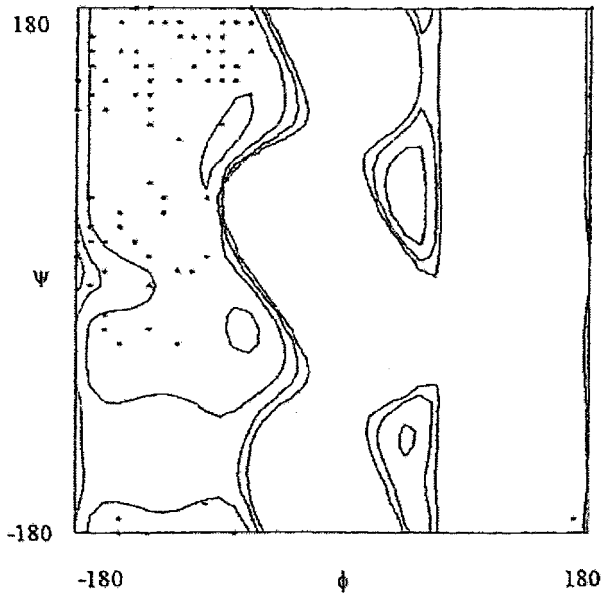


Figura 5 a

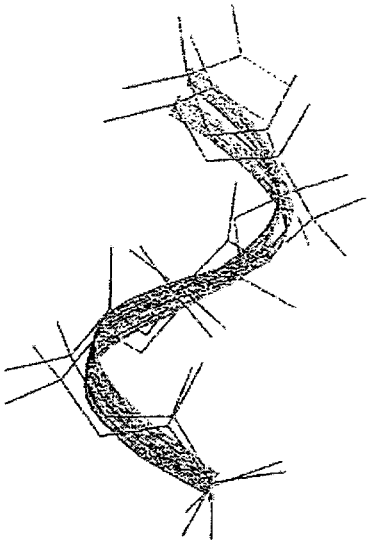


Figura 5 b

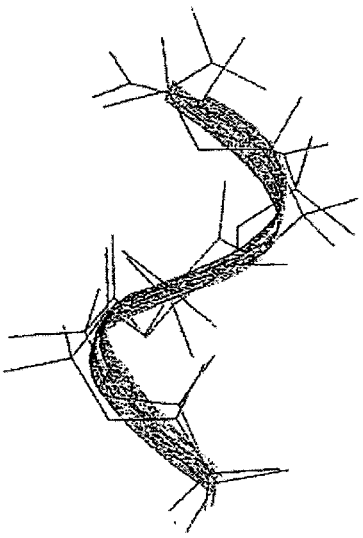


Figura 6

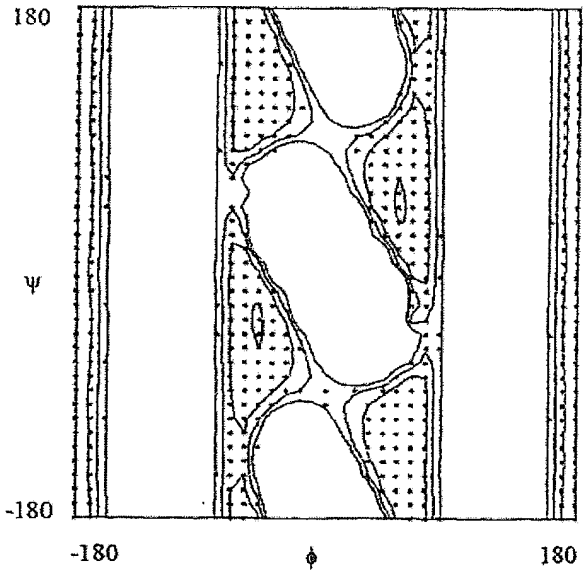


Figura 7

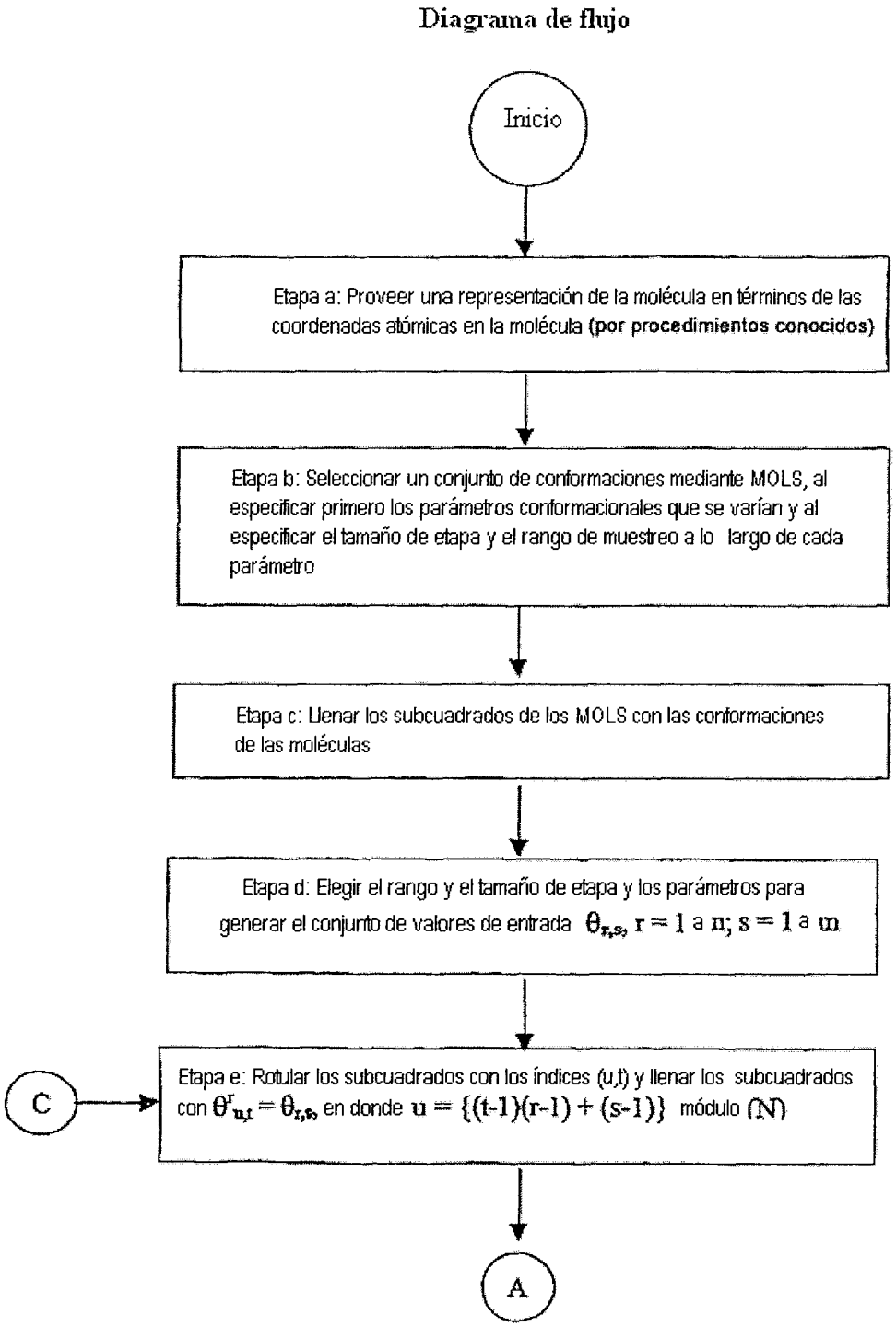

Diagrama de flujo
Inicio
Etapa a: Proveer una representación de la molécula en términos de las coordenadas atómicas en la molécula (por procedimientos conocidos)
Etapa b: Seleccionar un conjunto de conformaciones mediante MOLS, al especificar primero los parámetros conformacionales que se varían y al especificar el tamaño de etapa y el rango de muestreo a lo largo de cada parámetro
Etapa c: Llenar los subcuadrados de los MOLS con las conformaciones de las moléculas
Etapa d: Elegir el rango y el tamaño de etapa y los parámetros para generar el conjunto de valores de entrada $\theta_{r,s}$, r = 1 a n; s = 1 a m
C
Etapa e: Rotular los subcuadrados con los índices (u,t) y llenar los subcuadrados con $\theta^{r}_{u,t} = \theta_{r,s}$, en donde $u = \{(t-1)(r-1) + (s-1)\}$ módulo (N)
A

Figura 7 (cont.)

